DOCUMENT RESUME

ED 267 686 HE 019 154

AUTHOR	Walker, Gail; Kuchak, JoAnn
TITLE	Logistics and Sampling Plan for Task 2: 1979-1980 IRS Comparison Study. Quality Control Analysis of Selected Aspects of Programs Administered by the Bureau of Student Financial Assistance.
INSTITUTION	Applied Management Sciences, Inc., Silver Spring, Md.
SPONS AGENCY	Office of Student Financial Assistance (ED), Washington, DC.
PUB DATE	14 Jan 80
CONTRACT	300-79-0742
NOTE	31p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Students; Comparative Analysis; *Data Collection; *Disclosure; Eligibility; Error Patterns; *Federal Aid; *Financial Aid Applicants; Higher Education; Program Administration; *Quality Control; Records (Forms); Research Methodology; Sampling; Statistical Analysis; *Student Financial Aid
IDENTIFIERS	*Basic Educational Opportunity Grants; Internal Revenue Service; Misreporting

ABSTRACT

The type, number, and scope of errors on Basic Educational Opportunity Grant (BEOG) program applications were estimated in a replication of a 1976-1977 Internal Revenue Service (IRS) Comparison Study. Information reported on BEOG applicants and IRS income tax returns was compared for various categories of applicants. The study provides information on the impact of major quality control initiatives undertaken by BEOG during the 1978-1979 academic year, preaward validation (PAV), and the revisions to the application processing system edits. BEOG applicant files that were matched with IRS files consist of: 1979-1980 BEOG regular applicants with 1978 IRS file; 1979-1980 BEOG supplemental applicants with 1979 IRS file, and 1978-1979 BEOG quality control study group with 1977 IRS file. The PAV Study objectives center on assessing the relative efficacy of the validation process and the pre-established criteria for selecting applicants most likely to misreport information. Information is included on the logistics and sampling plans (e.g., selection of the preliminary sample, obtaining parents' social security number, finalizing the sample, IRS data extraction, and merging the IRS and BEOG data). Several procedural flowcharts are among the exhibits. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267686

G-170

Logistics and Sampling Plan for Task 2: 1979-1980 IRS Comparison Study

January 14, 1980

Prepared for:

Bureau of Student Financial Assistance
U.S. Office of Education
Department of Health, Education and Welfare

Contract No. 300-79-0742

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

NIE 119 154

APPLIED MANAGEMENT SCIENCES

APPLIED MANAGEMENT SCIENCES

962 Wayne Avenue • Suite 701 • Silver Spring, Maryland 20910
Telephone 301 585-8181

January 14, 1980

Ms. Sylvia Cowan
Project Officer
Division of Quality Assurance
Bureau of Student Financial Assistance
Room 4673, ROB #3
7th and D Streets, S.W.
Washington, D.C. 20202

Reference: Logistics and Sampling Plan for Task 2 of Contract No. 300-79-0742 "1979-80 IRS/BEOG Comparison Study" (AMS No. G-170)

Dear Ms. Cowan:

We are pleased to submit eight (8) copies of the above-referenced Logistics and Sampling Plan for distribution to IRS which specifies the procedures and research methodology for conducting Task 2 — 1979-80 IRS/BEOG Comparison Study. The Analysis Plan will be submitted under separate cover following a meeting with appropriate BSFA staff on preliminary specifications. This plan was prepared by Gail Walker under the supervision of JoAnn Kuchak.

We would like to schedule a meeting with you to discuss this plan prior to implementation. If you have any questions related to this plan, please do not hesitate to contact me or Ms. Kuchak.

Sincerely,

APPLIED MANAGEMENT SCIENCES, INC.

Todd S. Tucker, Ph.D.
President

TST/rcc
Enclosures

G-170

QUALITY CONTROL ANALYSIS OF SELECTED ASPECTS OF PROGRAMS ADMINISTERED BY THE BUREAU OF STUDENT FINANCIAL ASSISTANCE

This Logistics and Sampling plan is made pursuant to Contract No. 300-79-0742. The amount charged to the Department of Health, Education, and Welfare for the work resulting in this analysis plan is $11,514.00. The names of the persons employed by the contractor with managerial or professional responsibility for such work, or for the content of this analysis plan, are as follows:

Ms. JoAnn Kuchak, Project Director
Ms. Gail Walker, Senior Project Analyst

APPLIED MANAGEMENT SCIENCES, INC.
962 Wayne Avenue, Suite 701
Silver Spring, Maryland 20910

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Bureau of Student Financial Assistance

APPLIED MANAGEMENT SCIENCES

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1.1
	1.1 Overview	1.1
	1.2 1978-79 Pre-Award Validation Study and Corrections Study	1.2
	1.3 1978-79 Quality Control Study	1.4
	1.4 1976-77 IRS Comparison Study	1.5
	1.5 Summary	1.6
2	LOGISTICS PLAN	2.1
	2.1 Overview	2.1
	2.2 Selection of the Preliminary Sample	2.1
	2.3 Obtaining Parents Social Security Number	2.8
	2.4 Finalizing the Sample	2.9
	2.5 IRS Data Extraction	2.10
	2.6 Improvements to Match Rates	2.11
	2.7 Merging the IRS and BEOG Data and Performing the Analysis	2.12
	2.8 Summary	2.13
3	SAMPLING PLAN	3.1
	3.1 Overview	3.1
	3.2 Sampling Plan	3.3

LIST OF EXHIBITS

Exhibit	Title	Page
1.0	OVERVIEW OF METHODOLOGY AND AGENCY ROLES	1.7
2.0	PROCEDURAL FLOWCHART FOR PRELIMINARY SAMPLE SELECTION	2.2
2.1	PROCEDURAL FLOWCHART FOR SAMPLE SELECTION	2.3
2.2	PROCEDURAL FLOWCHART FOR FILE PREPARATION	2.4
2.3	PROCEDURAL FLOWCHART FOR IRS LOOK-UP	2.5
2.4	PROCEDURAL FLOWCHART FOR MERGE AND MERGE-YIELD	2.6
2.5	PROCEDURAL FLOWCHART FOR DATA ANALYSIS	2.7
3.0	SAMPLING FRAMEWORK FOR REGULAR AND SUPPLEMENTAL APPLICANTS	3.5

1

INTRODUCTION

1.1 Overview

The purpose of this project is to replicate the 1976-77 IRS Comparison Study in order to determine if any changes in rates of discrepancies or types of applicants with discrepancies have occurred during the ensuing three years. Therefore, the same basic procedures will be followed so that differences in the findings across years may be attributed to factors other than influences associated with methodological variations.

As in the prior effort, the intent of the 1979-80 IRS/BEOG Comparison Study will be:

- to provide an overall estimate of the type, number, and scope of errors entered on the entire population of Basic Grant applications; and,
- to identify specific categories of applicants who tend to misreport.

As such, the study will seek to determine the extent to which information reported on the Basic Grant application is similar to, or at variance with, IRS income tax returns for various categories of applicants.

The IRS study takes on additional importance this year because it will allow us to analyze the impact of major quality control initiatives undertaken by the Basic Grant Program during the 1978-79 academic year: pre-award validation (PAV) and the revisions to the application

processing system edits. Impact analyses of these new program changes were studied in the PAV and corrections analyses conducted by Applied Management Sciences in 1978-79. However, the criterion used to measure impact was based upon validation performed at the institutional level. Because the regulations requiring institutional validation did not become effective until February of 1978, and no data were systematically collected from institutions to determine whether validation had occurred, it is likely that the 1978-79 PAV and corrections analysis studies have overstated reporting accuracy. The IRS study will be able to provide better information for management decisions simply because the IRS represents a higher standard of verification quality than the applicant/recipient data that was used in the PAV and corrections analyses.

The BEOG applicant files that are to be matched with IRS files consist of the following:

- 1979-80 BEOG Regular Applicants with 1978 IRS file.
- 1979-80 BEOG Supplemental Applicants with 1979 IRS file. It is anticipated that the match rate for this group will be less than the other two groups as IRS will not have all 1979 income tax filers on the tape at the time of the match.
- 1978-79 BEOG Quality Control Study Group with 1977 IRS file.

Following are summaries of the three studies conducted during 1978-79 which will enhance the current IRS study.

1.2 Pre-Award Institution Validation Study and Corrections Study

The objectives of the PAV Study centered on assessing the relative efficacy of the validation process and of the pre-established criteria (PEC) for selecting applicants most likely to misreport information on

the Basic Grant application. Results of this study suggest that students who are selected for validation and who correct their applications after selection reduce their award potential more than applicants who submit unsolicited corrections. Analyses of the pre-established criteria used to select applicants for validation have shown that certain of the 1978-79 criteria (which have been revised for 1979-80) are very effective, while others are virtually ineffective. The corrections behavior of applicants who met the PEC (regardless of whether they were selected for validation) resulted in decreased award potential compared to that of randomly selected applicants. Since all applicants who meet PEC received edits it appears that the edits or the edits in concert with validation, are resulting in dramatic gains in reporting accuracy. Given these findings, the current IRS/BEOG Comparison Study will focus on determining the accuracy of data reported in response to the edits, as well as the accuracy of data validated by the financial aid officers.

The objectives of the Corrections Study were to address application processing behaviors and to review the corrections process to determine whether the new edit procedures instituted in 1978-79 elicited more accurate data than previous edit systems. The results of this study indicated that there was a sharp increase in the number of students submitting solicited corrections, especially to taxable and non-taxable income data fields. There was a dramatic reduction (26.6%) since 1976-77 in the proportion of unsolicited corrections made by students submitting applications before April 15. Solicited corrections were most often (63.4%) submitted by applicants who applied via the College Scholarship Service (CSS) applicants while unsolicited corrections were most often submitted (56.7%) by BEOG applicants who applied directly to BEOG. Students in universities and other 4-year schools submitted more solicited and mixed corrections than students in less than 4-year schools.

The current IRS/BEOG Comparison Study will enable BSFA to see which groups are making the greatest number of corrections to adjusted gross income, and household size. The effectiveness of edits dealing with financial data will be carefully studied. The current study will also determine which edits seem to result in a change in the student's eligibility index.

1.3 1978-79 Quality Control Study

In 1978, the Division of Quality Assurance engaged Westat, Inc. and Macro Systems Inc. to conduct a quality control study of the Basic Grant program to ascertain the extent of errors being made by Basic Grant program applicants as well as the impact of these errors upon total program expenditures. This study examined the following three questions:

- What is the overall level of error in the program?
- What are the principal types of errors and what are the reasons for these errors?
- What management actions can be taken to reduce the level of error?

The study sample was comprised of 225 postsecondary institutions, stratified by type, control and geographic location and a total of 2,309 dependent and independent students and their parents who were representative of approximately 1.35 million recipients.

Data were collected through personal interviews with the financial aid officer at each institution, personal interviews with students and parents, record abstracts of student data from school files, Student Eligibility Report (SER) transaction history, published school materials and data from the Program Information Management System (PIMS) recipient file.

Results of this study identified characteristics of error prone applicants. It was found that all of the financial variables (adjusted gross income, social security benefits, home value and investment debt) contributed to overawards. As the value of these variables increased, the likelihood of an overaward increased. Applicants with large household sizes rarely received an overaward. In situations where underawards were received, errors in computing non-taxable income and net investment assets were most frequently discovered. Dependent students received more underawards than independent students.

The current BEOG/IRS Comparison Study will allow some of the variables which contribute to overawards and underawards to be examined closely. Due to the fact that the Quality Control Study was conducted on 1978-79 applications, it is now possible to conduct detailed analyses as data are available regarding actual BEOG award payments. The 1979-80 IRS study provides two opportunities for assessing the field data collection effort:

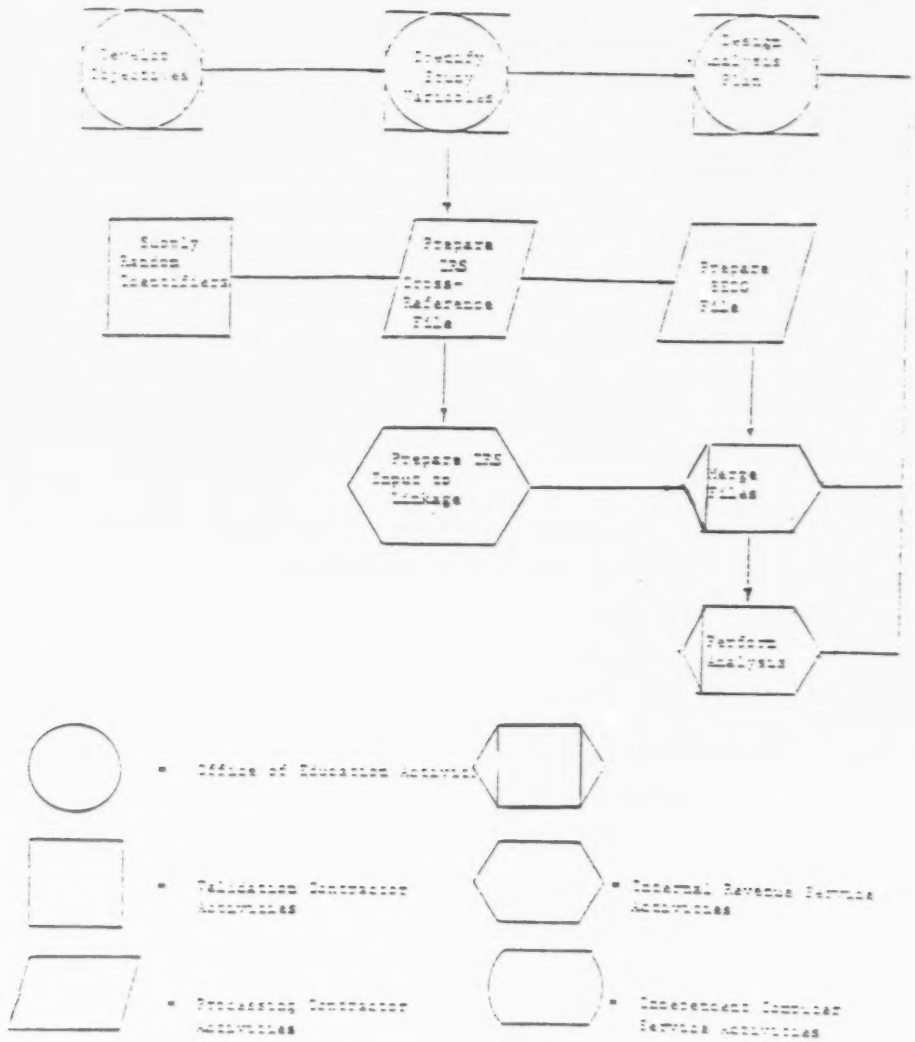
1. It will allow comparisons to be made between the BEOG-QC verified values and the IRS-reported values; and
2. It will provide an opportunity to study QC nonrespondents to determine whether these individuals exhibit greater rates of misreporting than the respondents, and therefore the extent to which the QC findings may be biased.

This information will be useful to OE in order to obtain more accurate estimates of program-wide applicant misreporting. It will also help to determine whether the field data collection approach results in obtaining "true" copies of Federal tax documents, and, depending on how the QC data were collected, whether alternative documents should be accepted for validation purposes.

1.4 1976-77 IRS Comparison Study

In 1978, BSFA contracted with Applied Management Sciences to conduct an IRS Comparison Study of 1976-77 Basic Grant applicants, which was a replication of the 1974-75 study. This study analyzed the effectiveness of several BEOG quality control procedures and investigated issues concerning income earned by dependent applicants. A slight decrease in the overall accuracy of applicant reporting was noticed compared with 1974-75 data, especially in the case of higher income applicants. Results of this study pointed to a need to determine whether or not the newer and tightened edit procedures implemented in 1978-79 effect increases in the accuracy of income reporting. The current study will investigate this question.

EXHIBIT 1.0: OVERVIEW OF METHODOLOGY AND AGENCY ROLES



BEST COPY AVAILABLE

2

LOGISTICS PLAN

(Logistics, Security and Confidentiality)

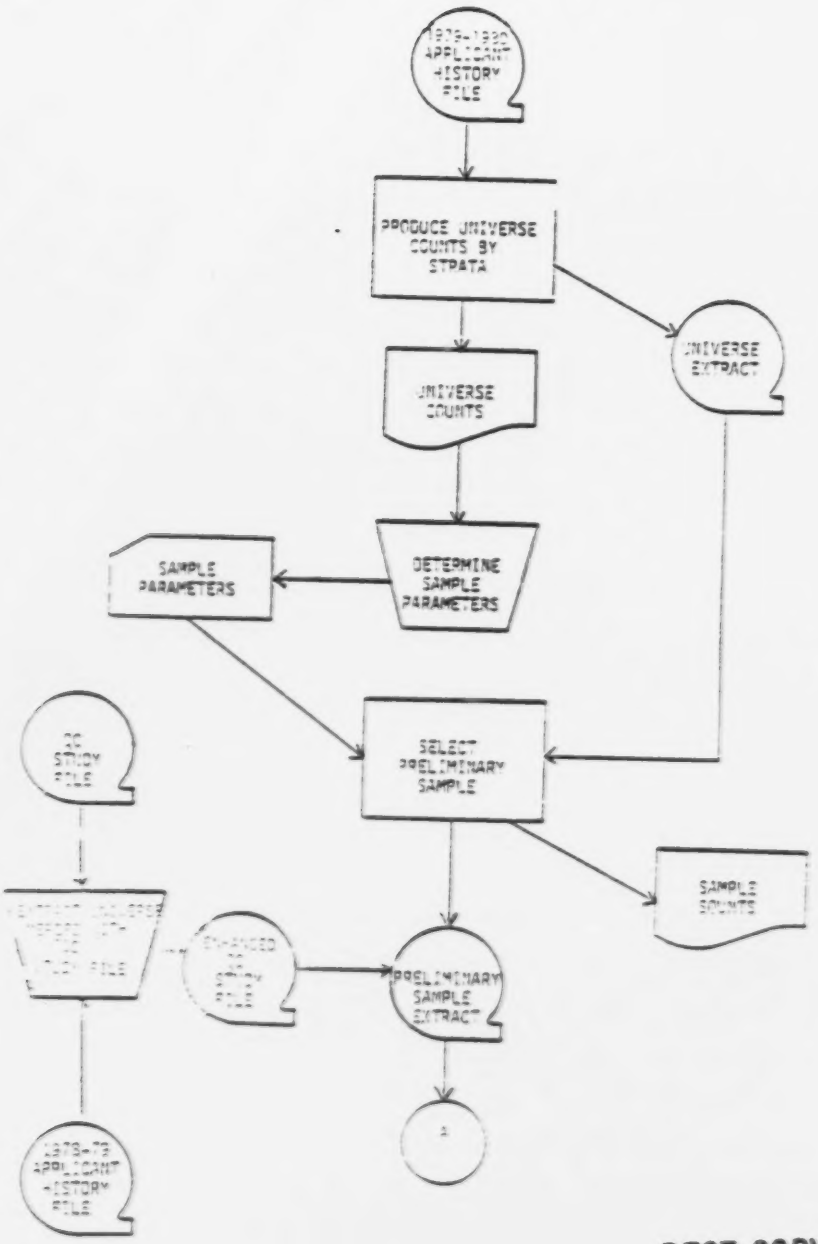
2.1 Overview

The purpose of this chapter is to present a plan for the major logistical and data processing procedures for the IRS/BEOG component of this project that will fulfill the requirements of the Federal Privacy Act, as well as legislation governing the IRS. The discussion follows the procedural flowcharts presented in Exhibits 2.0-2.5. This plan is similar to the successful plans used in the 1974-75 and 1976-77 IRS/BEOG Comparison Studies. In both of these studies there was no time at which any of the agencies involved in the BEOG/IRS tape match had access to BEOG individual student data as well as individual IRS data.

2.2 Selection of the Preliminary Sample

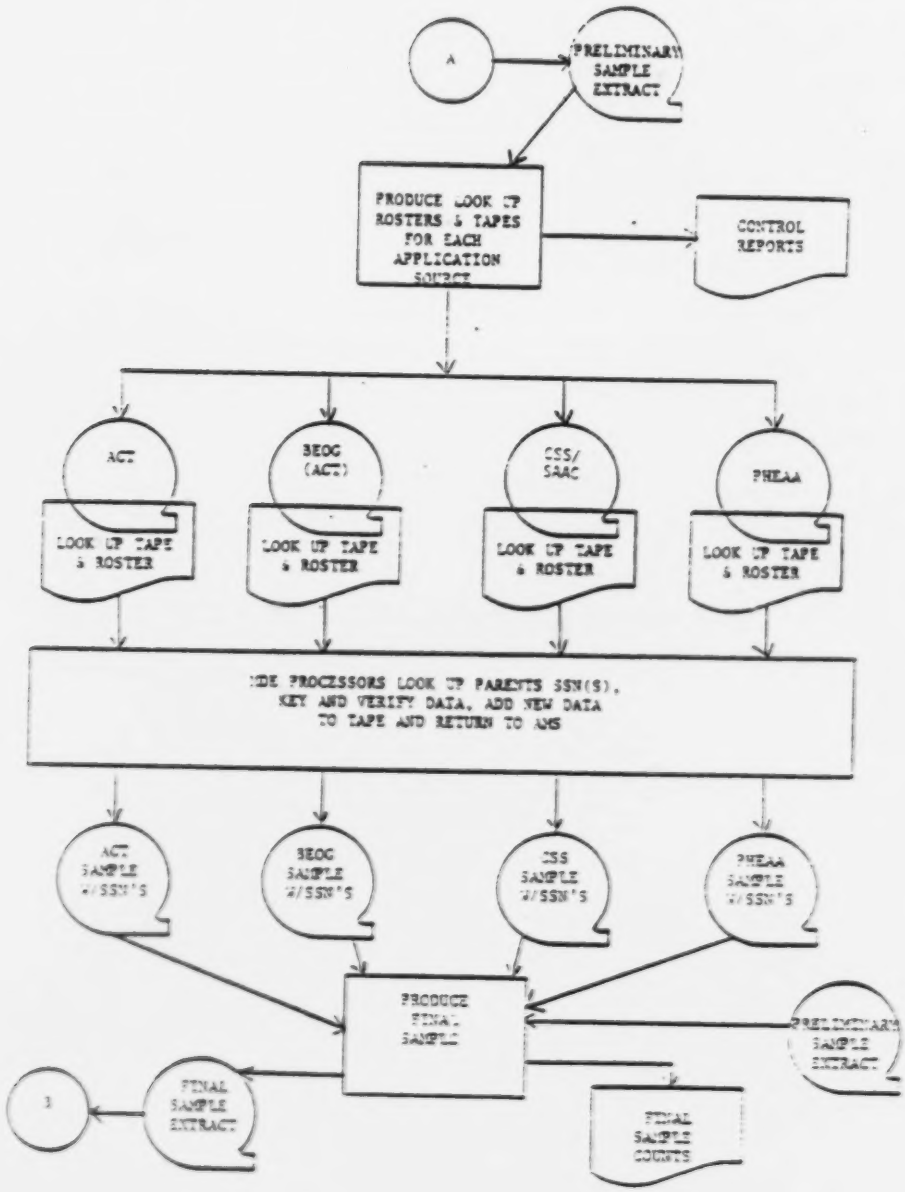
The first step in selecting the preliminary sample is to pass the 1979-80 BEOG Applicant History File to obtain universe counts for the strata to be used to define the sample. The program to carry out this activity will also match all applicants who were included in the 1978-1979 BEOG Quality Control Study. It is assumed that the Task Monitor for this project component will supply Applied Management Sciences with a copy of this tape along with the appropriate documentation. The 1978-79 Applicant History file will be used for QC applicants. The QC applicants will be flagged and put into separate strata. The program will produce an abbreviated extract file of the potential universe of applicants and a report detailing the number of applicants in each of the strata, as well as the Quality Control data. An overview of the logistics plan is contained in Exhibit 2.5.

EXHIBIT 2.0: PROCEDURAL FLOWCHART FOR PRELIMINARY SAMPLE SELECTION



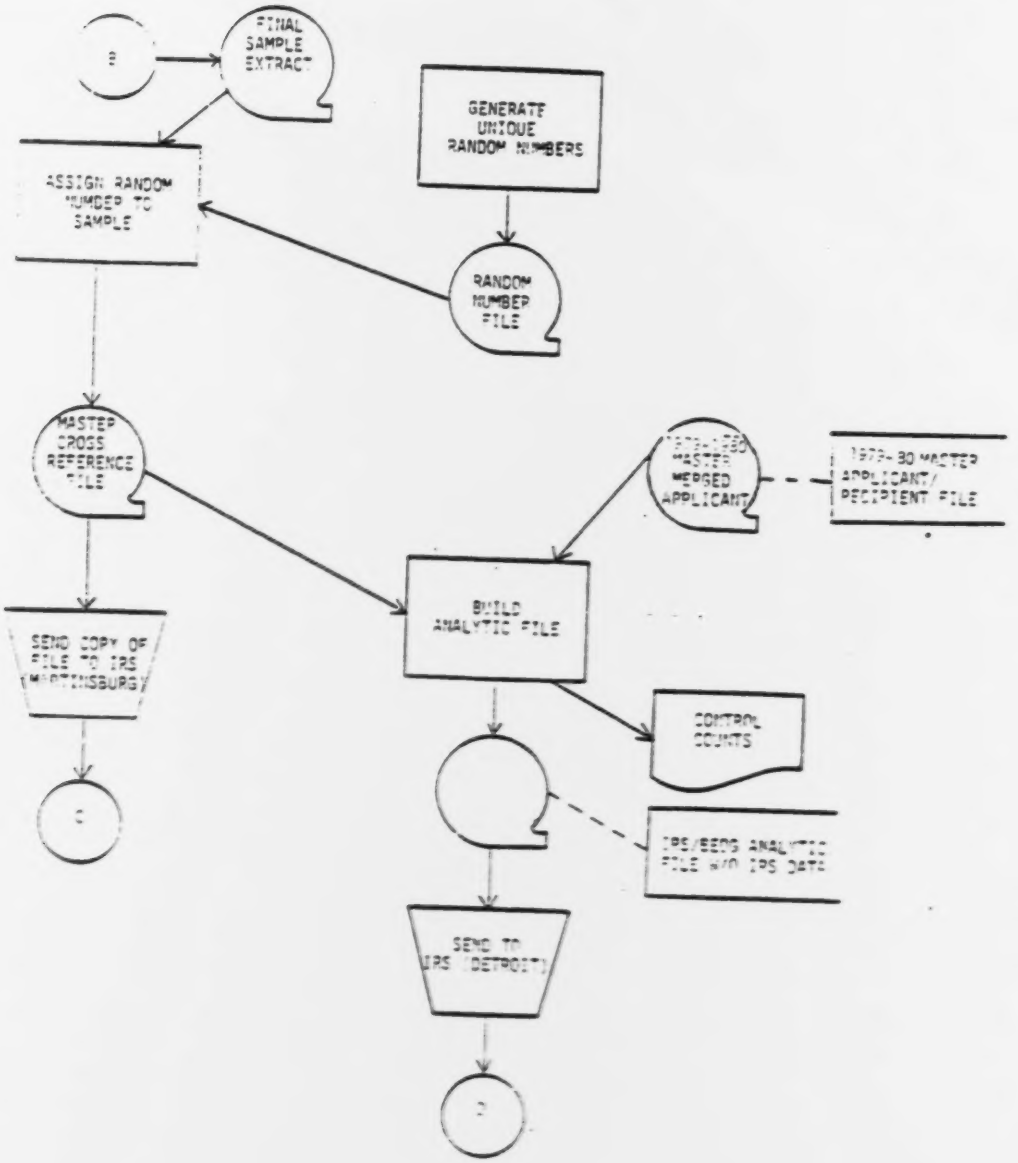
BEST COPY AVAILABLE

EXHIBIT 2.1: PROCEDURAL FLOWCHART FOR SAMPLE SELECTION



BEST COPY AVAILABLE

EHNIBIT 2.2: PROCEDURAL FLOWCHART FOR FILE PREPARATION


FINAL SAMPLE EXTRACT
GENERATE UNIQUE RANDOM NUMBERS
ASSIGN RANDOM NUMBER TO SAMPLE
RANDOM NUMBER FILE
MASTER CROSS REFERENCE FILE
1979-80 MASTER APPLICANT/ RECIPIENT FILE
SEND COPY OF FILE TO IRS (MARTINSBURG)
BUILD ANALYTIC FILE
CONTROL COUNTS
SEND TO IRS (DETROIT)

BEST COPY AVAILABLE

EXHIBIT 2.3: PROCEDURAL FLOWCHART FOR IRS LOOK-UP

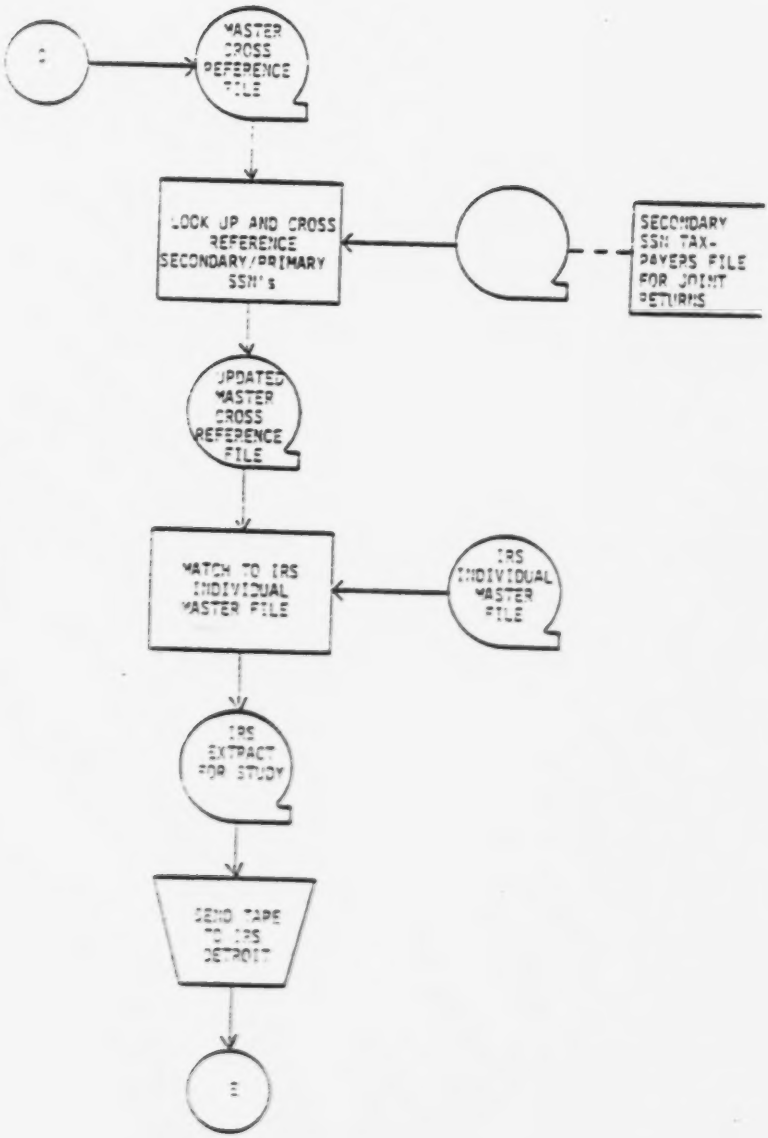

MASTER CROSS REFERENCE FILE
LOOK UP AND CROSS REFERENCE SECONDARY/PRIMARY SSN's
SECONDARY SSN TAX-PAYERS FILE FOR JOINT RETURNS
UPDATED MASTER CROSS REFERENCE FILE
MATCH TO IRS INDIVIDUAL MASTER FILE
IRS INDIVIDUAL MASTER FILE
IRS EXTRACT FOR STUDY
SEND TAPE TO IRS DETROIT

EXHIBIT 2.4: PROCEDURAL FLOWCHART FOR MERGE AND MERGE-YIELD

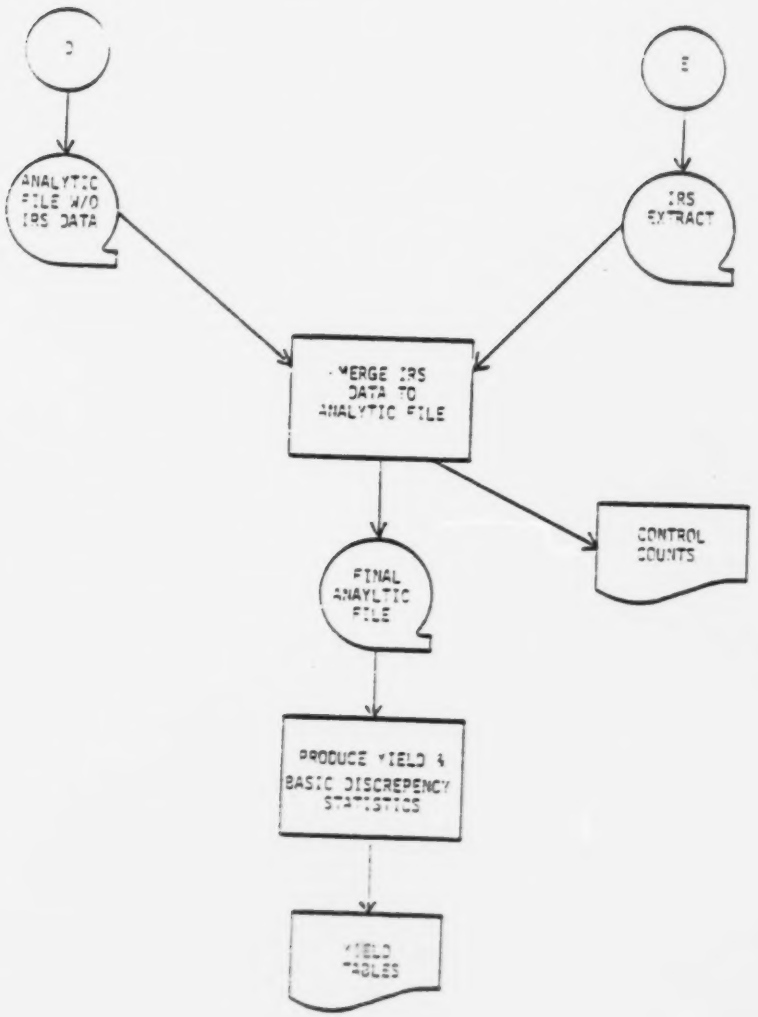
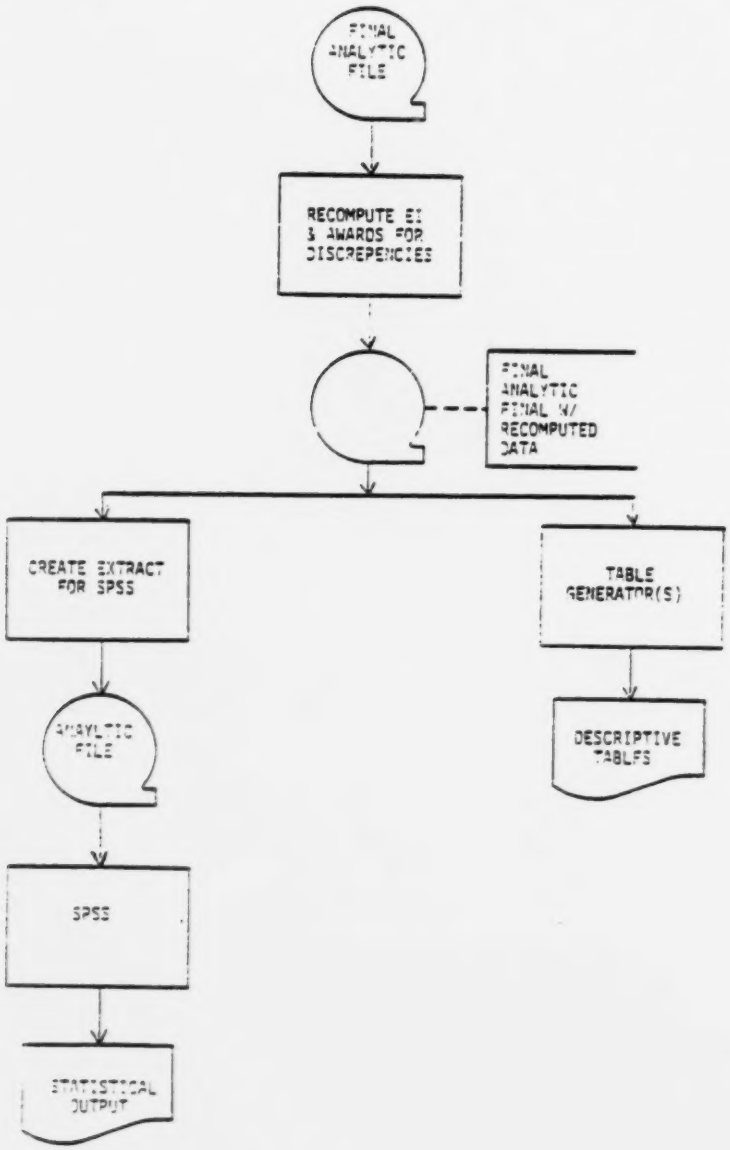


EXHIBIT 2.5: PROCEDURAL FLOWCHART FOR DATA ANALYSIS



It should be noted that we are using recipient data for the Quality Control group in the analysis, since actual payment data will be available. Preliminary payment data will be used for the 1979-80 BEOG applicants.

The technical staff and consulting statistician will be provided with the output of the first program and they will be responsible for determining the number of applicants to be selected from each of the strata. A total sample of approximately 400,000 is planned. These figures will be keyed and used as input to the second program which will select the sample. It should be mentioned that this is only the preliminary sample, and, that we will select applicants for this sample at a higher rate than we expect to obtain in the final sample. This is done because a number of dependent applicants do not provide the parent's SSN on the application. Therefore, those who do not provide this information cannot be matched with IRS data. The sample selection program will also produce an extract tape of all applicants chosen for the preliminary sample.

2.3 Obtaining Parent's Social Security Numbers

Students may apply to the Basic Educational Opportunity Grant Program by filling out one of several available application forms (the BEOG form or the form provided by the Multiple Data Entry sources (MDE) listed below:

1. College Scholarship Service (CSS)
2. American College Testing Program (ACT)
3. Pennsylvania Higher Education Assistance Agency (PHEAA)
4. Student Aid Application for California (SAAC)

Since the BEOG application processing system does not contain the social security number for the parent, it will be necessary to obtain this information from the hardcopy application. With the Multiple Data Entry System for the 1979-80 year it will be necessary to make arrangements with all five proccessors to obtain these social security numbers as well as the last name of the parent. Subcontracts are being arranged at this time.

Applied Management Sciences will produce a hardcopy roster (in a sequence to be specified by each of the MDE processors) and a computer tape containing the same information. The processors have been asked to look up the parent's SSNs for all applicants (including independents) and append this information to the tape. The MDE processors may conduct this activity in one of two ways, depending upon their own system. Some have this information on a computer file and therefore a match is all that is necessary. Others will have to look this information up on the application, transcribe it, key it and then match it to the file. For those requiring the second approach, Applied Management Sciences will require that the processor perform a validation check to assure quality in the look-up and keying processes. The completed files will be returned to Applied Management Sciences for further processing.

2.4 Finalizing the Sample

Once the MDE processors have completed their look-up actitivies, the sample will be finalized. This will be accomplished in several steps. First, Applied Management Sciences staff will prepare a file of unique random indentifiers. This tape will be used to code all of the files for matching purposes, since the real identifiers cannot be used because of confidentiality reasons. This file, along with the tapes from MDE processors, the preliminary sample extract file, and the 1978-79 merged applicant recipient file will be forwarded to the independent computer center, American Management Systems, for final processing. Applied Management Sciences will prepare and test all of the necessary software for this phase, but our personnel will not run the jobs in production mode or handle any of the data. This is necessary to assure that no concerned party conducting the study has access to the random identifiers and the live data at the same time.

The independent processor is the American Management Systems Computer Center, which has performed this same service for the past two BEOG/IRS studies that Applied Management Sciences has conducted for OE. They will be responsible for running the program that merges the MDE tapes with the preliminary sample extract and the random number file to create the final

sample file. (A by-product of this program is a count by stratum indicating how many records were dropped due to the absence of the parent's SSN.) Once the program has run successfully, they will forward a copy of the tape to the appropriate IRS official for further processing in Martinsburg, West Virginia. Note that this file will only contain the information necessary for the IRS look-up and will contain no BEOG data.

The second major program that Applied Management Sciences will run is the creation of the preliminary analytic file, which will be created using the merged applicant/recipient file. The output file will contain no personal identifiers except the random identification code assigned to the case. Once this file has been created, it will be forwarded to the IRS officials at the Detroit administrative data center.

2.5 IRS Data Extraction

The 1976-77 IRS study required that IRS perform a two-stage process to extract data from their master file of individual taxpayers. The same two-step procedure will be employed for this study. The first step is to access their secondary SSN file. This is a file which contains the secondary social security number for all forms which were filed as a joint married return. It is possible for the BEOG applicant to record the secondary taxpayers' SSNs where IRS has the return filed under the primary SSN. In this instance, we would not obtain a match. The process of first accessing the secondary SSN cross-reference file allows IRS to substitute the correct SSN before accessing the master file, thus ensuring a proper match. Once the SSNs have been reconciled a match will be made to the full Individual Master File and an extract of the requested data will be made for those records that matched successfully. IRS will apply the hierarchy for determining matches described in the following section. If the record resulted in a partial match or a mismatch, an indicator is set on the output file to let us know that this condition occurred. Once all processing is complete, the extract file will be forwarded to the IRS Detroit facility.

2.6 Suggested Improvements to Match Rates

The match rate, or number of successful BEOG-IRS linkages attained, is the key to achieving representation and generalizability. The social security number (SSN) of the independent applicant (and spouse) and the dependent applicant's parent(s) is the primary linkage medium.

There are several features of the BEOG and IRS processing system which cause non-matches to occur. One reason is SSN reporting errors (innocent or intentional). The major reason, however, is that in the case of dependent applicants, the BEOG program requests, but does not require parent social security numbers. As a result, some applications contain no parent identifier and it is not possible, within the limitations of the study design, to obtain parent social security numbers. This, in turn, reduces match rates and introduces possible biases into the findings. In past IRS studies, a match was considered valid only if both the SSN and first two letters of the last name matched on both IRS and BEOG records. This was done to prevent mismatches due to keypunch or reporting errors. However, this restriction also prevents valid matches from occurring when an individual has changed his/her last name. In the 1976-77 IRS Comparison Study, 25.6 percent of the applications resulted in a non-match. Given that significant numbers of individuals change marital status each year, this is a major factor that may affect match rates. It is also possible that unintentional transpositions occur when reporting social security numbers. This will be investigated in any feasible way to increase match rates due to transpositions. Therefore, to improve the match rate without risking erroneous matches, we recommend that a match be considered valid if one of the following conditions are met:

1. SSN and first two letters of last name match exactly; or
2. SSN matches; first two letters of the last name do not match; selected components of the address do match such as zipcode;
3. SSN on BEOG files are transpositions of SSNs on the IRS file or the SSNs are inconsistent by one digit and (a) name matches and (b) selected components of the address match.

Through the application of this decision hierarchy, we are confident of being able to increase the rates of valid matches.

2.7 Merging the IRS and BEOG Data and Performing the Analyses

All data processing activities for the remainder of the project will be run using the IRS Detroit computer facility which can be accessed by remote job entry stations in IRS' Washington office.

When running at the IRS computer center the following protocol will be used: first, the job will be prepared by the Applied Management Sciences personnel and forwarded to an IRS official who will submit the job; second, when the job has been run the output will be examined by the same IRS personnel; third, the job control log will be examined by the Applied Management Sciences programmer to make sure the job completed successfully; fourth, the remainder of the output from the job will be forwarded to the IRS coordinator for the study and will be examined to ensure no identification codes are present or cells representing less than 10 individuals appear on the output; and lastly, the output will be forwarded to Applied Management Sciences for analysis. It should be noted that the programs prepared by Applied Management Sciences will automatically blank out cells where less than 10 observations are present. Additionally, any percents or other statistics associated with that cell will also be blanked out.

Two phases of analysis will be conducted. The first phase will involve the merge of the two sources of data and the production of the yield statistics. The merge program will join the two data sets together and form the final analytic file. A report will be produced from the merge program displaying the count of true matches by strata. The yield statistics will provide further elaboration of the matching and non-matching populations and will also provide basic tabulations of discrepancies. This output will be used by the technical staff to prepare the interim report.

Phase II processing will involve the recomputation of the Eligibility Index and the Award for those applicant records found to be in error. Two types of analyses will be performed during Phase II, bivariate contingency tables will be produced and either discriminant analysis or stepwise regression analysis will be used to improve the criteria for identifying error prone cases. This output, in conjunction with the Phase I output will be used to prepare the final report.

2.8 Summary

The procedures outlined in the discussion above have been derived from the procedures that have been used previously, with much success, for the two previous IRS studies conducted by Applied Management Sciences for OE. They provide OE with a full analytic capability and also prevent any unauthorized access to individual data. These procedures were approved by IRS, and unless regulations have changed very recently, will satisfy their confidentiality requirements. In summary, this plan incorporates the following security and confidentiality features:

- No exchange of individually indentifiable data between OE and IRS.
- Removal of personal identifiers from all analytic files.
- Limitation on minimum cell size on tabular output to 10.
- Security precautions for file storage.
- Approval of programming and analysis plans by IRS and OE.

Under this task we will work closely with OE to finalize the preliminary strategies described herein. After OE approval is received, a copy will be forwarded to IRS for review and approval by their Disclosure Staff.

3

SAMPLING PLAN

3.1 Overview

The purpose of this chapter is to define the study population and the methods that will be employed to select the sample. The following list presents the applicant population subgroups that we believe will need to be studied in order to respond to these objectives:

Objective	Study Groups
1. Assess overall scope of misreporting	Total Applicant Population 1a. Total Regular applicants 1a1. Regular eligible filers 1a2. Regular eligible filers who were paid 1a3. All ineligibles 1a4. Rejected Regular filers 1b. All supplemental filers 1b1. Supplemental eligible filers who were paid 1b2. Supplemental eligible filers 1b3. Supplemental ineligible filers 1b4. Supplemental rejected filers
2. Assess impact of misreporting	2. Same as 1

Objective	Study Groups
3. Assess trends in reporting accuracy	Compare data for the years 1976-77, 1978-79, and 1979-80 for the following groups: 3a. Total regular eligibles and ineligibles 3b. Total regular eligibles 3c. Total supplementals 3d. Supplemental eligibles
4. Compare PAV results with IRS	4. PAV sample applicants
5. Compare QC results with IRS	5. All QC sample applicants
6. Assess QC response bias	6. QC sample applicants who did not produce documentation
7. Assess effectiveness of edits	7a. Applicants who received certain edits 7b. Applicants who did not receive certain edits 7c. PEC applicants who received certain edits 7d. PEC applicants who did not receive certain edits
8. Assess the effectiveness of the PEC	8a. PAV sample applicants 8b. PEC group 8c. Random group
9. Improve criteria for identifying error-prone cases	9. Regular eligible applicants who were paid

If OE concurs with the objectives and the corresponding subgroups, then our strategy for sampling will have to assure adequate representation of each group. This will be accomplished by using these groups to define strata in a random, stratfied sampling approach. A total sample size of 400,000 with a sampling error not to exceed five percent at the 95 percent confidence level for the total sample will be easy to achieve with such a large sample size. In fact, this total sample size will result in virtually negligible error at the 99 percent confidence level.

In the past IRS studies, eligibility status, dependency status, income level, and presence/absence of corrections served as key stratification variables. In this study we want to place more emphasis on the edits and corrections, plus we want to study the PEC selection criteria, to evaluate the effectiveness of institution validation, and to evaluate the effectiveness of the QC procedures. Therefore, we must ensure adequate representation for all of these groups.

3.2 Sampling Plan

With a sample size of 400,000, it will not be necessary to use all of the variables described above, plus income and dependency status for stratification purposes for all groups. For the PAV criterion, PAV random and QC groups, we will only stratify by major subgroups to assure that we have adequate representation for the major analytic issues. Therefore, our current thinking is to recommend the following strata and approximate sample sizes to result in the total of 400,000.

TOTAL SAMPLE		400,000
Total Regular Applicants		370,000
Eligibles		270,000
• with comments, but not in QC or PAV	101,600	
• without comments, but not in QC or PAV	100,000	
• Total PAV	66,000	
.. met criteria	50,000	
.. random	16,000	
• QC sample (1978-1979)	2,400	
Ineligibles		50,000
• with comments	30,000	
• without comments	20,000	
Rejected		50,000
Total Supplemental Applicants		30,000
Eligibles		15,000
Ineligibles		5,000
Rejected		10,000

As is evident from this design, we have placed most emphasis on eligibles and applicants who received comments. The smallest applicant subgroups studied (i.e., those who misreport and those who received

specific comment (types)) will come from these groups. Least emphasis is given to supplementals, about whom we are only interested in overall findings. Rejected and ineligible applicants will also be studied with regard to comments, so the sample size for these groups is somewhat larger. We believe that this design will fulfill the study needs; however, we will review our rationale and the proposed sample sizes with OE before finalizing our design.

Exhibit 3.0 displays the sampling framework for regular and supplemental applicants.

EXHIBIT 3.0: SAMPLING FRAMEWORK FOR REGULAR AND SUPPLEMENTAL APPLICANTS

Dependency Status	REGULAR APPLICANTS					SUPPLEMENTAL APPLICANTS		
	Eligible Applicants		Ineligible Applicants		Rejected	Eligible	Ineligible	Rejected
	with comments, but not in QC or PAV	without comments, but not in QC or PAV	with comments	without comments				
Dependent								
[illegible]								
Independent								
[illegible]								

BEST COPY AVAILABLE